

papers. The school I have chosen is somewhat special, and belongs to a social science Academia.

[illegible]

Ministry of Economic Affairs for various research, training, and creative and technology-related activities; the National Science Council (NSC) for all academic programs relating to industrial development; and the National Science Council (NSC) for all academic programs relating to

mation

WAIS, WIDE AREA INFORMATION SERVERS, IS A CLIENT-SERVER SYSTEM WHICH ALLOWS INFORMATION RETRIEVAL TO BE DONE OVER A SET OF HETEROGENEOUS INFORMATION COLLECTIONS. IT IS RAPIDLY BECOMING A DE-FACTO STANDARD ON THE INTERNET, WITH INCREASING NUMBERS OF ORGANIZATIONS MAKING INFORMATION AVAILABLE ELECTRONICALLY BY OPERATING A WAIS SERVER. BREWSTER KAHLE HEADED THE WAIS PROJECT AT THINKING MACHINES AND THEN, IN JULY '92, SPLIT OFF TO FORM WAIS INCORPORATED, A SOFTWARE AND CONSULTING COMPANY. WHEN I SPOKE WITH KAHLE I WAS INTERESTED IN FINDING OUT WHY CERTAIN USER INTERFACE AND ARCHITECTURE CHOICES WERE MADE AND IN WHAT HE SAW AS THE GOALS AND FUTURE OF THE WAIS PROJECT.

Interview of Brewster Kahle

USER INTERFACE

I've always been surprised that WAIS uses the natural language query technique because there is so much evidence that it often causes the naive user to attribute too much intelligence to the software. Have you run into this problem at all?

Well, it's interesting to watch the query's that come in. Sometimes people overstate what the computer can do, but what people are extremely good at is figuring out what they can get away with. Children can size up a substitute teacher in about five minutes. It's the same thing with our users, they can figure out what the server does and what it doesn't do very easily. What natural language allows us to do is grow. It doesn't lock us into a particular query language that will die after a year. It gives us a lot of flexibility.

One of the WAIS documents mentions that relevance feedback didn't end up being used much because users found it conceptually confusing. Have you made any progress on this issue?

No, we have some ideas but nothing concrete. Relevance feedback starts to pay off when you have a really good server and when you've got an information collection of over a gigabyte. And on the Internet we have neither. There are freeware servers out there which aren't very good, that don't do a very good job with relevance feedback. And most of the collections of information are relatively small. Boolean searches, seed word searches, work pretty well for up to a couple hundred megabytes. What relevance feedback does is give us reason to believe that WAIS will scale to extremely large servers.

What sort of future user interface techniques are you looking at? Are you exploring the intelligent agent concept at all?

Oh absolutely. I don't tend to use the word 'agent' much because it's anthropomorphic; it was great to get funding with but it doesn't necessarily lead to good engineering practices.

But automating the requesting process so you are presented with timely information is something we're exploring. I also think personal newspapers make a lot of sense.

But really what we're trying to do here is to create a critical mass of servers, so that as people build new user interfaces they have an infrastructure to plug into. And we're seeing that happen; there are now twelve different user interfaces publicly distributed on the Internet. And on the server side we're seeing people clip SQL databases onto the back-end of the WAIS protocol, and USGS has made a server that understands latitude and longitude queries and will send you the appropriate map. So using the protocol as an infrastructure and changing the server or the client so that it fits a particular task is what we're excited about.

Right now most of the WAIS user interfaces around are really just windows on the protocol. User interfaces need to grow more towards making existing applications WAIS enabled. Wouldn't it be neat if your word processor was WAIS enabled? You could just mark a word and ask what it means and the application would go off to the dictionary server or the encyclopedia server. Or WAIS enabled e-mail that would look up where an address is. Gopher is an example of an application which uses WAIS, and I think it's a better application because of it.

ARCHITECTURE

It's interesting that you went with the Z39.50 information retrieval protocol, you were really one of the first products to use the standard. What influenced that decision? It seems that going with a new and untested protocol poses some real dangers.

When we started the project we really wanted to use an existing protocol because otherwise it would have been seen as Thinking Machine's protocol, or Apple's protocol, and we wouldn't have been able to get other people involved. So we looked around at the existing standards, but all of them were terrible. Then I talked with some of the people on the Z39.50 committee and asked, "if we were to come on with all our corporate pals, could we change the protocol fundamentally and radically?" And they basically said yes. So that's what we did. We did end up extending the protocol some, so that it would allow multimedia and really large documents, and these changes will be reflected in the new version of Z39.50.

Do you think you lost anything by having to try to stick within the framework of the protocol?

No, I think we have only gained. Not because it's a particular standard, but just because it is a standard. What we're trying to do is bypass the proprietary protocol period — and it's risky. If we screw up, if the world splits into a million competing variants (as UNIX has), it leaves us very vulnerable to a proprietary solution. But what the big corporations said, the Apple's and Dow Jones's, is that we need to have an open standard because the most crucial thing is to achieve critical mass. So what exactly is in the protocol matters a whole lot less than that it *is* an open protocol.

But doesn't this make your own position tenuous? With an open protocol there's no reason why anyone needs to go to you for server or client software.

Absolutely. That's the way it should work: level the playing field, and then win. And frankly, we're too small, Apple is too small, Dow Jones is too small, to dominate the market. The market for this stuff is just way too big. So lots of people making servers that extend to lots of other markets is great. Thinking Machines has developed a high-end server, and I hope Apple develops a low-end server. It's the filling out of the market that will win over the users.

What do you see as the next steps for the WAIS architecture?

Well the architecture seems to be doing O.K., but the protocol needs to be stretched in a few different ways. Right now there are limitations on the size of document lists; that needs to be changed. We also need mechanisms for server forwarding, for dealing with heterogeneous networks.

ACCESS

Don't systems like WAIS increase the barriers to access of information for the poor and those who don't have access to computers?

The dissemination of this technology is happening at a phenomenal clip. Take the introduction of the printing press in 1452. By the year 1500 there were a million books. That's pretty amazing, but it still was only the rich and well educated who had these books. And it stayed that way for about another

hundred years. So the dissemination of that technology took 150 years. The Internet, on the other hand, is doubling every seven months. The dissemination has gotten to the point where in a few short years large numbers of people have gained access to the net.

I think the key thing to ask about this technology is "what can allow cheaper use?" The WAIS technology is built to run over any kind of communication system, not just the Internet. This was a big battle within the Z39.50 committee, they just wanted to support OSI. And we, the commercial players, needed it to work over ISDN, over modems, and over X.25. We weren't just rich boys in government labs. We want to get to lots and lots of users. We're already seeing growth in K-12 and in the smaller colleges. We estimate that about 20,000 people have used WAIS so far, from 28 countries. There are 350 servers available right now, and it's doubling every six months.

What sort of effect do you see the NREN having on WAIS?

Tremendous proliferation of the net. The largest holding block for us is not search technology, not copyright law, not the publishers — it's that we need a reliable digital infrastructure. Having to have a Ph.D. and know what DTR means to use a modem doesn't qualify. America has a great voice system, but we've been slow in developing a digital one. I see NREN as a mechanism to spur the United States towards a reliable digital infrastructure. And what that will mean is that more people will be able to use WAIS.

What will WAIS's policy be on charging for information and royalties?

Well, publishers are very interested in coming up with new ways to distribute their information. WAIS is built on the centralized publishing model so you can continue to have control over access. We are not prescriptive like Xanadu is in trying to establish a payment policy. The information provider is free to charge any way he wants to. So you could have the first 30 days be free, or you could do pay-per-search, or pay-per-retrieval. WAIS can support any method.

So you're basically following the MIT/X dictum of "mechanism, not policy."

Exactly. We see this as the only way to make WAIS durable.

CODA

What's your personal goal in all of this? What do you hope to see?

I want to live in the future we're creating. That's what fundamentally motivates me. I think the key is not so much that people will be able to find more information, but that more people will be able to put information out. More people will be able to be publishers. A lot of our satisfaction with our jobs and all is being able to do things that other people like and use. And if we could put more people in a position to be able to do things that other people like and use we'll have a happier—and more efficient—society. So what I'm excited about is finding the large numbers of people who are not publishers now, but who want to be.

But who's going to be reading all of this?

Well maybe publishing is the wrong word because it makes people think that it's all high quality. I think of it as bridging the gap somewhere between a dinner party and a magazine. Electronic publishing is so much cheaper than hard copy that it allows you to write for a much smaller audience. It will foster a whole lot of different communities.

ACCESS TO WAIS

WAIS is accessible over the Internet and can be used to query many different free databases, including the CIA World Factbook and the Columbia Law School library catalog.

There are several ways to get to WAIS. You can telnet to it or use the Macintosh or DOS client versions. These clients are freely available through anonymous ftp.

Telnet access: `telnet quake.think.com`
login: `wais`
password: `your_username`

DOS Windows: `ftp to <ftp.oit.unc.edu> and get:`
`/pub/wais/UNC/Windows`

Macintosh: `ftp to <think.com> and get:`
`/wais/WAISStation-0-63.sit.hqx`

The USENET groups `alt.wais` and `comp.infosystems.wais` can provide additional information on WAIS issues. A moderated discussion list is available from:

`wais-discussion-request@think.com.`

WHEN IT WAS PROCLAIMED THAT THE
LIBRARY CONTAINED ALL BOOKS,
THE FIRST IMPRESSION WAS ONE
OF EXTRAVAGANT HAPPINESS. ALL
MEN FELT THEMSELVES TO BE THE
MASTERS OF AN INTACT AND SECRET
TREASURE. THERE WAS NO
PERSONAL OR WORLD PROBLEM
WHOSE ELOQUENT SOLUTION DID
NOT EXIST IN SOME HEXAGON. THE
UNIVERSE WAS JUSTIFIED, THE
UNIVERSES SUDDENLY UNSURPED THE
UNLIMITED DIMENSIONS OF HOPE.